# Profiling Essential Professional Skills of Chief Data Officers Through Topical Modeling Algorithms

*Full Paper*

**Wei Dai**
University of Arkansas at Little Rock
wxdai@ualr.edu

**Ningning Wu**
University of Arkansas at Little Rock
nxwu@ualr.edu

## Abstract

Today enterprises are increasingly dependent on data to keep their business competitive and successful. To better harness values of data, more and more organizations are establishing Chief Data Officer (CDO) position. The professional skills of CDOs are rather diverse because CDOs are expected to undertake a variety of roles in their companies including enterprise data architect, data quality and governance manager, business strategy leader, business regulation compliance officer, etc. CDO is an emerging research field, few studies have been done on CDO. This paper tries to profile what are the key professional skills and education background that current CDOs have by studying their resumes on LinkedIn using topic modeling technique. This work is a step forward towards understanding the roles of CDOs in organizations and what are the professional skills and experience they may need have in order to undertake their responsibilities of managing data and realizing its true values for their organizations.

**Keywords:**

Chief Data Officers, CDO, social network analysis, topic modeling

## Introduction

In the era of big data, companies are increasingly dependent on data to make sure that their business are competitive and successful. Data becomes a critical asset of organizations. Data harness affects almost all business processes, from customer experience to risk managements, operational efficiency, targeted marketing, and much more.

To better harness values of data, more and more companies are establishing Chief Data Officer (CDO) position. Gartner predicts that 50% of all companies in regulated industries will have a CDO by 2017. CDOs help manage data as an organization asset. They are playing a strategic role in organizations due to the importance of data analytics, data science, and big data. According to Gartner, the CDO is a senior executive who bears responsibilities for the firm's enterprise wide data and information strategy, governance, policy development, and effective exploitation. CDO's roles will combine accountability and responsibility for information protection and privacy, information governance, data quality and data life cycle management, along with the exploitation of data assets to create business value (McCall, 2015).

As CDOs are an emerging new breed of executive, few studies have been done in the field. A number of articles discussed the roles and evolution of CDOs, but there is no common standard on those roles because they vary from company to company. Many questions are still open about CDOs. For example, what are the key skills CDOs should have, what are the preferred professional background/experiences they need to have, what are their career paths? There are very limited research papers on CDO. A paper by Lee *et al* proposed a three-dimensional cubic framework that describes the roles of CDO based on a longitudinal study of 10 years that was comprised of informal case studies, detailed iterative interviews, and structured surveys (Lee 2014). The framework describes roles of CDO in three dimensions: collaboration direction, data space, and value impact, and it identifies eight different CDO role profiles, which are coordinator CDO, report CDO, architect CDO, ambassador CDO, analyst CDO, marketer CDO, developer CDO, and experimenter CDO. CDOs are expected to be capable of undertaking any of those roles when needed. As more companies will establish a CDO position, it comes with, from the employer's perspective, the questions such as what makes the breed of CDOs, what kind of persons are the best candidates for a CDO, what are the preferred skills and background CDOs should have, etc. The similar questions might be asked by people who want to plan their career path towards CDO, or by colleges who want to project the career path for their graduates.

Existing studies on CDO are conducted primarily from business perspectives and focus on understanding why CDOs are needed in organizations, what are the roles of CDO, and how CDOs may help organizations in their business. This paper tried to study CDOs from a different angle – CDOs skill sets and their educational background. In the last decade, professional social networks (PSNs) have become a popular tool for people to share their professional life and to establish business relationships and connections. These networks contain rich information about people's profile such as education and training, skills, job transitions, business connections, etc. PSNs are also widely used by companies to share their job openings and attract qualified people for those positions, by job seekers to find appropriate jobs, and by job placement professionals to find best talent for specific jobs.

In fact, there is the increasing popularity of using data mining techniques on PSNs data to address human resources management problems. This paper attempted to identify the common skill sets of a group of CDOs by analyzing their resumes on LinkedIn, a professional social network, using topic modeling technique. Based on the identified skill sets, we hoped to gain a better understanding of what are the key roles played by current CDOs in organizations. In addition, the findings of the paper could be used by organizations to identify the people of desired skills for a CDO position.

The paper analyzed resumes of 621 people enrolling in a CDO group on LinkedIn. Of 621 people, 34 are CDOs based on their official job titles, and 228 are de-facto CDOs (i.e., their companies haven't had an official CDO position yet but they have assumed the roles and responsibilities of CDO). Among 228 CDOs, 201 have education information on their resumes, and 184 have skill information. The paper applied Non-negative Matrix Factorization, a topic modeling method, on the skill information of CDOs' to identify five key skill sets including general management, business intelligence, business and strategy, data management, and system, business, financial, and policy analysis. The paper also discussed the relationship between the skill sets and CDO role profiles proposed by a cubic three-dimensional CDO framework (Y. Lee *et al* 2014). To the best of our knowledge, the paper is the first attempt to describe CDOs' key skills quantitatively from their professional resumes on PSN using data mining techniques.

## Related Work

This section reviews the related work in two areas: CDO research and data analytics on a person's profile using social network data.

### *Research on CDO*

CDO is an emerging research field. Limited research has been done in the field. Lee *et al* proposed a three-dimensional cubic framework to describe the role of the CDO (Y. Lee *et al* 2014). The three dimensions are collaboration direction, data space, and value impact. The framework is based on a 10-year study that consists of informal case studies, iterative interviews and structured surveys. The framework provides a guide for organizations as they analyze the need for a CDO and will enable them to determine the most appropriate profile for their CDOs now and in the future.

It has been a trend that CEOs are increasingly adding the CDO role to their management teams to tackle the big business issues that come with data. There are several articles that discuss CDO's roles and the challenges they face from the business aspects. Berkooz discussed the ways that CDO can get their companies to collect clean data by using five levers to show why spending time on data quality matters (Berkooz 2017). Steele argued that the two distinct realms currently covered by CDOs, data strategy and data governance will only become more important with time going (Steele 2015). In addition to the private sectors, an increasing number of local and federal governments have created the position of CDO to lead their efforts toward data-driven government. Wiseman studied CDOs in government by documenting selected current practices, including advice shared by existing government CDOs, observations by the author, and analysis from government technology and analytics expert (Wiseman 2017). The study aims to help new entrants to understand the challenges they will face and to learn the best practices from pioneer CDOs. The PWC article (Savelloni, Vazquez, Jonson, Yakowitz, and Nair, 2015) discussed the key CDO roles and challenges in financial sector and identified the key actions that will help CDO on the path to success.

### *Social Network Analysis*

Analyzing a person's profile based on their information on social networks is an interesting topic and has attracted large attention in the research field. Due to the popularity and proliferation of social network applications, people especially young generations tend to use those applications as a quick and convenient way to share their personal and professional life with others. There are many applications of data analytics on social network data. Consumer opinion mining, user profiling and behavior analysis, friendship and acquaintance network analysis, marketing, and business intelligence are just few examples of social network analysis. In this paper, we will focus on research on data analytics on professional social networks.

There are quite a few research efforts on user's professional profile and occupation analysis from social networks. For example, Preotiuc-Pietro *et al* inferred occupational class for a user based on user profiles and social contents from a single source –Twitter (Preotiuc-Pietro, Lampos, and Aletras 2016). Filatova *et al* presented a methodology for identifying three-level hierarchy of biological activities and then used the occupation-related activities as features for classifying the occupation of a person (Filatova and Prager 2012). Liu *et al* proposed a multi-source multi-task learning model that uses information from multiple social networks to describe a user and characterize properties of his or her career (Liu, Zhang, Nie, Yan, and Rosenblum 2016). Xu *et al* proposed a professional similarity measure framework that takes into account the temporal information in a person's career path to assess the career similarity between two individuals (Xu, Li, Gupta, Bugdayci, and Bhasin, 2014). The framework can be used to predict the career path of a person or to identify the candidates whose professional profile matches the requirement of a particular job. Xu *et al* utilized online professional networks to identify talent circle in job transition networks. Based on the identified talent circle, the paper developed a talent exchange prediction method for talent recommendation (Xu, Yu, Yang, Xiong, and Zhu 2016).

## Methodology

Professional social networks(PSNs) provide a useful platform for professionals to share their profiles and establish business relationships. Most PSNs provide templates for their members to publish their resumes. Common features in a resume include education, employment and experiences, skills and endorsement, membership of professional groups, connections with other members, etc.

CDO, emerging as a key leader in the organization, is playing a critical role in helping the organization to utilize information to identify market opportunities, increase shareholder value, protect data security and privacy, meet regulatory demands, etc. In general, CDO' job responsibilities cover business domain, technical fields, and team management. Furthermore, different industries require different skills, resulting in CDO being an executive with diverse background and skills. So it is hard to characterize their skills and experiences.

This research aims to gain a deeper understanding of CDO by analyzing their professional profiles shared on LinkedIn. It focuses on identifying CDO's professional skills and educational background based on social network analysis and topical modeling techniques. Our methodology involves the following steps:

1. Collect data from a CDO group on LinkedIn.
2. Clean data by removing dirty data and null values, correcting spelling errors, and standardizing terms with different names.
3. Extract professional skills based on the topic modeling technique.
4. Discuss our findings and their relationships with an existing CDO cubic framework

## *Data Collection*

Information of more than 450 million members in over 200 countries and territories is maintained by LinkedIn in 2016. Most members post their resumes including educational background, working experience, and professional skills on LinkedIn. Table 1 shows the general data schema of online resumes. Figure 2 shows featured skills and endorsement of a sample resume on LinkedIn.

According to Fawley (2013), endorsements on LinkedIn are equal to unofficial reference letters. In this paper, we focused on professional skills of CDO. We assumed that most professional skills endorsed by LinkedIn members are accurate data.

| Sections | Context |
|---|---|
| Brief Profile | name, current job title and company, city and region, the number of connections |
| Experience | personal experience |
| Education | Education degrees, majors, or training background |
| Featured Skills & Endorsements | Professional skills endorsed by others |
| Accomplishments | Publications, patents, and awards |
| Following | Following companies, groups, and schools |

**Table 1. Data Schema of LinkedIn**

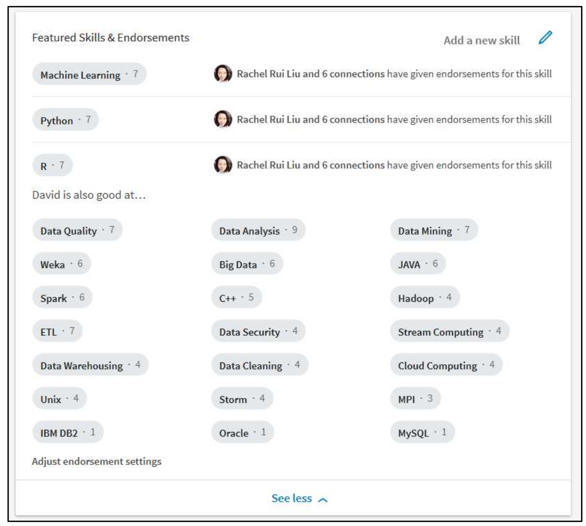


**Figure 2. Featured Skills and Endorsements**

The paper selected a LinkedIn group named the *MIT Chief Data Officer (CDO) & Information Quality Symposium* as the focus of our study. The group is managed by Dr. Richard Wang and has 621 members as of February 28, 2017. Its members are either CDOs or have interests in CDO and information quality. The members' resumes were used for this study.

## *Data Cleansing and Selection*

There are several major data quality problems in the resumes, including non-English words, null values, nonstandard words, and typos. The data cleansing process focuses on the following tasks:

- Identifying non-English words: the most popular non-English vocabularies are French words. For example, 'Stratégie IT' means 'Strategy IT'; 'Gestion de projet' means 'project management'. Using ASCII codes, resumes including the foreign language have been identified and ignored for further analysis.
- Dealing with Null values: professional skills of some members are left as null values if they didn't input any skills on LinkedIn. These resumes have been identified and ignored for further analysis.
- Correcting misspelled words: typos are a common problem in online resumes. For example, "data mangement" instead of "data management". Python Natural Language Toolkit (NLTK) is used to identify and correct the misspelled words (Bird 2009).
- Data standardization: There are many variations of a same degree program shown in the resumes. For example, MBA, M.B.A, or Master of Business Administration. These variations cause problems in data analysis. We standardized the degree program names so that each degree program has a single unified name.

In the CDO group on LinkedIn, not every member is CDO. After manually reviewing job titles, experience, and educational background, we identified 262 members who are either CDO or de facto CDO (i.e., their companies haven't had an official CDO position yet but they have assumed the roles and responsibilities of CDO). The job titles of de facto CDOs are quite diverse, including 'Global Data and Information Operations Manager', 'Principal Data Systems Engineer', and 'Data Management & Analytics Leader'. After data cleansing process on 262 CDO resumes, 184 of them have professional skills, and 201 of them have education information. Thus these two groups of resumes are used in this research.

Table 2 lists professional skills and education majors of 2 randomly selected CDOs; Table 3 lists the information of 2 non-CDOs.

| | | |
|---|---|---|
| $CDO_1$ | Skills | Data Quality, Data Analysis, Data Management, Data Integration, Business Intelligence, Databases, Data Warehousing, SQL, Data Mining, ETL, Data Modeling, Requirements Analysis, Business Analysis, Information Management, Master Data Management, Leadership, Systems Analysis, Analytics, Big Data, Java, Analysis, Microsoft SQL Server, Oracle, Enterprise Software, Statistics, SDLC, Programming, Visio, Database Design, SAS, Software Project Management, Software Development, Integration, Enterprise Architecture, Requirements Gathering, Agile Methodologies, Business Process, Data Cleaning, UML, MS Project, Business Objects, SAS programming, Data Migration, IT Strategy, Database Marketing, Access, Informatica, Big data, Extract, Transform, Load (ETL) |
| | Majors | Business Administration, Management Information Systems |
| $CDO_2$ | Skills | Business Intelligence, Enterprise Architecture, Data Warehousing, Business Analysis, Master Data Management, Data Modeling, Project Management, Requirements Analysis, SDLC, IT Strategy, Data Architecture, Data Governance, Consulting, Program Management, Integration, Strategy, Vendor Management, Metadata Management, Data Management, Data Mining, Information Management, ERP, Data Integration, Business Development, Requirements Gathering, Inventory Management, Inventory Control, Data Warehouse Architecture, ETL |
| | Majors | Industrial Engineering, Industrial and Systems Engineering |

**Table 2. Sampling professional skills of CDO members**

| | | |
|---|---|---|
| Non-$CDO_1$ | Skills | Analytics, Leadership, Communication Coaching, Data Science, Coaching, Consulting, Recruiting, CRM, Business Development, Public Speaking, Marketing, Research, Social Networking, Sales, Market Research, Time Management, Teaching, Editing, Publishing, Advertising, Academic and Professional Research, Interviews, Conflict Management, Outlook, Information Analysis and Synthesis, Social Media, Public Relations, Qualitative Research, Microsoft Word, International Business Development, PowerPoint, Debate, Executive Search, Management Consulting, Management, Human Resources, Customer Relationship Management (CRM), Interviewing |
| | Majors | MBA, Psychology |
| Non-$CDO_2$ | Skills | Enterprise Software, CRM, Business Intelligence, New Business Development, SaaS, Selling, Big Data, Cloud Computing, Sales, Solution Selling, Go-to-market Strategy, Professional Services, Sales Management, Pre-sales, E-commerce, Analytics |
| | Majors | Null Value (Does not input data) |

**Table 3. Sampling professional skills of non-CDO members**

## *Data Analysis*

Topic modeling technique has been widely used in text mining for identifying key topics in a collection a documents. This paper used Non-negative Matrix Factorization (Lee 2001) to extract "skill" topics from resumes. Non-negative Matrix Factorization (NMF) is an unsupervised family of algorithms that simultaneously perform dimension reduction and clustering. It approximates a nonnegative matrix **A** by the product of two low-rank nonnegative matrices **W** and **H,** where **W** is basic vectors, and **H** is coefficient matrix. In terms of topic modeling applications, given a collection of documents and the corresponding term-document matrix **A,** NMF computes **W** and **H** to approximate **A,** where **W** represents the topics (clusters) in the data, and **H** represents the membership weights for documents relative to each topic (cluster). Since it gives semantically meaningful result, NMF has been widely used as a document clustering method and as a topic modeling method. Due to its property, NMF performs well in topic modeling for short text documents (Xu 2003 and Yan 2013).

Another popular topic modeling algorithm is Latent Dirichlet Allocation algorithm(LDA) (Blei 2003). The LDA is a three-layer hierarchical Bayesian model that extracts mixtures words with probabilities to represent documents.

The paper assumes the skills of each resume as a document, and aims to find the key professional skillsets (identified as topics) in a collection of CDO resumes. Due to the fact that the description of a skill often contains much fewer words than a typical sentence does, NMF method was selected in this study. In addition, LDA was used so that we can compare the NMF results with it.

The scikit-learn library, an open source Python library for data mining and data analysis, is used for implementing NMF and LDA in our analysis.

## Results and Analysis

### *Extracting Essential skills*

For both NMF and LDA algorithms, we set parameters as follows: the number of topics is five; the max of n-gram is two. We have tested different settings for the parameter number of topics, and found that when the number of topics is five, the result is more coherent and easier to interpret. Table 4 and Table 5 show results of both algorithms. For each topic, the first 20 top phrases are displayed. We manually assigned skill set name to each topic in both tables based on their top words.

| TopicID (skillset name) | Top Phases |
|---|---|
| Topic_1<br>(Management) | Management \| Data Management \| Program Management \| Program \| Vendor Management \| Vendor \| Change Management \| Change \| Master Data \| Master \| Risk Management \| Risk \| Consulting \| Management Consulting \| Information Management \| Information \| Product Management \| Product \| Team Management \| Portfolio Management |
| Topic_2<br>(Business Intelligence) | Intelligence \| Business Intelligence \| Business \| Business Development \| Development \| Business Strategy \| Business Analysis \| Business Transformation \| Business Analytics \| Transformation \| Intelligence Tools \| Tools \| Objects \| Business Objects \| Business Requirements \| Artificial Intelligence \| Artificial \| Business Planning \| Planning \| New Business |
| Topic_3<br>(Business and Strategy) | Strategy \| Business Strategy \| Marketing Strategy \| Business \| Marketing \| Go-To-Market Strategy \| Go-To-Market \| Digital Strategy \| Digital \| Information \| Data Strategy \| Commercial \| Strategy Development \| Language \| Language Processing \| XML \| Latex \| Launch \| Lead \| Lead Generation |
| Topic_4<br>(Data management -- analysis, architect, governance, integration, modeling, and quality) | Data \| Data Warehousing \| Warehousing \| Data Management \| Data Modeling \| Modeling \| Governance \| Data Governance \| Big Data \| Big \| Data Analysis \| Integration \| Data Quality \| Master \| Master Data \| Quality \| Data Integration \| Data Mining \| Mining \| Data Architecture |
| Topic_5<br>(System, business, financial and policy analysis) | Analysis \| Business Analysis \| Business \| Data Analysis \| Requirements Analysis \| Requirements \| Data \| Financial Analysis \| Systems Analysis \| Systems \| Financial \| Requirements Gathering \| Gathering \| Competitive Analysis \| Competitive \| Business Requirements \| Business Development \| Business Strategy \| Development \| Policy Analysis |

**Table 4. Professional skills based on the NMF algorithm**

| TopicID (skillset name) | Top Phases |
|---|---|
| Topic_1 (Business and strategy) | Strategy \| Analytics \| Business \| Marketing \| Business Analytics \| Business Strategy \| Services \| ERP \| Digital \| SAS \| Operations \| Web \| Professional \| Predictive \| Marketing Strategy \| Insurance \| Professional Services \| Health \| Credit \| Predictive Analytics |
| Topic_2 (Data management – integration, quality) | Data \| Integration \| Quality \| Software \| Data Quality \| Enterprise \| Enterprise Software \| Databases \| Data Integration \| Outsourcing \| Engineering \| Science \| Statistics \| Sigma \| Finance \| Supply \| Supply Chain \| Chain \| Documentation \| Data Science |
| Topic_3 (Business Analysis) | Business \| Process \| Analysis \| Development \| Improvement \| Business Process \| Business Analysis \| Process Improvement \| SDLC \| Software Development \| Software \| Business Development \| Sales \| Systems \| Public \| Transformation \| Speaking \| Business Transformation \| Public Speaking \| Objects |
| Topic_4 (Business Intelligence) | Intelligence \| Business \| Business Intelligence \| CRM\| Financial \| Design \| Oracle \| Database \| Database Design \| Agile \| Agile Methodologies \| Methodologies \| Customer \| Services \| Financial Services \| Presales \| Building \| Performance \| Team \| Team Building |
| Topic_5 (Information Technology and management) | Information \| Management \| Leadership \| SQL \| Change \| Healthcare \| Change Management \| Technology \| Information Technology \| Sap \| Information Management \| Microsoft \| Healthcare Information \| Training \| Informatica \| Disaster \| Recovery \| Disaster Recovery \| Information Architecture \| Technical |

**Table 5. Professional skills based on the LDA algorithm**

From the top phases of each topic, we can see that NMF is more effective in extracting topics (skill sets) since these phases in each of its topics are more coherent than those identified by LDA. Table 6 lists topic index along with their corresponding manually assigned skillset names of each method. The common skills identified by both methods are business intelligence, and business and strategy. Data management is the skill identified by both methods; however, NMF identified more sub-skills under Data management. Another common sub-skill identified by both is business analysis. The unique skill and sub-skill identified by NMF include management, data modeling and governance, system analysis and financial analysis. The unique skills and sub-skills identified by LDA are information technology and management, and business development. Overall, NMF captures more unique and coherent skills than LDA.

| Topic index | NMF | LDA |
|---|---|---|
| 1 | Management | Business and strategy |
| 2 | Business intelligence | Data management –integration, quality |
| 3 | Business and strategy | Business analysis and development |
| 4 | Data management -- analysis, architect, governance, integration, modeling, and quality | Business intelligence |
| 5 | System, business, financial, and policy analysis | Information technology and management |

**Table 6. Skill Sets Comparison**

Figure 2 shows word clouds of topic phrases by NMF. We consider that the top words, which make each topic, may identify the major skills that most CDOs have. The most prominent words should indicate the most common skills that CDOs have. These skills may indicate the mostly commonly used skills of CDO, or may indicate the most demanding skills of CDO. Our future research will investigate what are these skills meant to CDO.



**Figure 2. Displaying skills based on the NMF** **Figure 3. Education Majors of CDOs**

### *CDO's Educational Background Analysis*

Educational background is important to CDOs because people learn knowledge, practice skills, and establish social network in schools. Based on the group of 201 CDOs whose resumes have education information, on average each CDO has received two degrees. 16.9% of them have a Ph.D. degree or doctoral degree; 62.2% have at least one master degree. In terms of degree concentration, 47.8 of them have a computer science, computer engineering, or information science degree; 31.8% of them have a MBA degree; 16.0% of them have a business, finance, or economics degree.

### *Professional skills and a CDO Cubic-Framework*

Next, we discuss the relationship between professional skills of CDO and a CDO cubic-framework proposed by Lee *et a*l. The CDO cubic-framework contains three dimensions including collaboration direction (inwards and outwards), data space (conventional data and big data), and value impact (service and strategy). We rearranged professional skills based on CDO roles, as shown in Table 7, in an effort to map the skills to the framework.

| CDO Dimension | Professional Skills |
|---|---|
| Collaboration Direction | Management, Program Management, Vendor Management, Risk Management, Management Consulting, Team Management, Lead, |
| Data Space | Data Management, Master Data, Information Management, Business Intelligence, Business Analysis, Artificial Intelligence, Intelligence Tools, Data Strategy, Digital Strategy, XML, Data Warehousing, Data Modeling, Information, Data Quality, Data Mining, Data Architecture, System Analysis, Data Governance |
| Value Impact | Business Development, Business Strategy, Business Objects, Business Planning, Business Transformation, Marketing Strategy, Go-To-Market Strategy, Commercial, Strategy Development, Requirement Analysis, Financial Analysis, Product Management, \| Portfolio Management |

**Table 7. Rearranging Professional skills based on the NMF algorithm**

## Conclusion and Future Work

Even though there are many ways to be a CDO, having essential professional skills and education background is an important factor for a successful CDO. This paper analyzed the resumes of over 200 CDOs of a LinkedIn group. It aimed to gain a deep understanding of CDO by identifying the key skills that those CDOs have and the characteristics of their educational background. A topic modeling technique was used to identify the major skill sets and key skills from their resumes. It also discussed the relationship between professional skills and the CDO cubic-framework by mapping the skills to CDO roles.

Our future work contains two directions. First, we will include more features in our study, including working experience, sequence and transitions in one's career path, business connections, and memberships, to build a better profile of CDOs. Second, we will perform the qualitative study to evaluate and validate our findings.

## REFERENCES

Berkooz, G. 2017. "How Chief Data Officers Can Get Their Companies to Collect Clean Data," Harvard business Review, Feb. 16, 2017.

Bird, S., Klein, E., and Loper, E. 2009. "Natural language processing with Python: analyzing text with the natural language toolkit," O'Reilly Media, Inc.

Bishop, C. M. 2006. "Pattern recognition," Machine Learning 128 (2006): 1-58.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent dirichlet allocation," Journal of Machine Learning Research 3. Jan (2003): 993-1022.

Coelho, L. P., and Richert, W. 2015. "Building machine learning systems with Python," Packt Publishing Ltd, 2015.

Fawley, N. E. 2013. "LinkedIn as An Information Source for Human Resources, Competitive Intelligence." Online Searcher (2013): 31.

Filatova, E., and Prager, J. 2012. "Occupation Inference through Detection and Classification of Biographical Activities". Data & Knowledge Engineering 76:39-57, 2012.

Lee, Y., Madnick, S., Wang, R., Wang, F., and Zhang, H. 2014. "A Cubic Framework for the Chief Data Officer: Succeeding in a world of big data," MIS Quarterly Executive, 2014.

Lee, B., Riche, N. H., Karlson, A. K., and Carpendale, M.S. T. 2010. "Sparkclouds: Visualizing Trends in Tag Clouds." IEEE transactions on visualization and computer graphics 16.6 (2010): 1182-1189.

Liu, Y., Zhang, L., Nie, L., Yan, Y., and Rosenblum, D. S. 2011. "Fortune Teller: Predicting Your Career Path," in *Proceedings* of the 13th AAI Conference on Artificial Intelligence, Beld, Slovenia.

McCall, T. 2015. "Understanding the Chief Data Officer Role," http://www.gartner.com/smarterwithgartner/understanding-the-chief-data-officer-role/. Feb. 18, 2015.

Preotiuc-Pietro, D., Lampos, V., and Aletras, N. 2015. "An Analysis of the User Occupational Class Through Twitter Content," in Proceedings of the Conference of the Association for Computational Linguistics, pages 1754-1764.

Stanganelli, J. 2017. "Chief Data Officer Role Continues to Evolve," *Enterprise Apps Today,* Aug. 2016.
Steele, J. 2015. "Understanding the Chief Data Officer," http://www.oreilly.com/data/free/, Feb. 2015.
Wiseman, J. "Lessons from Leading CDOs: A Framework for Better Civic Analytics," Ash Center for Democratic Governance and Innovation at Harvard University's Kennedy School.
Lee, D. D., and Seung, H. S. 2000. "Algorithms for Non-negative Matrix Factorization," in *Proceedings of* Advances in Neural Information Processing Systems.
Savelloni, M., Vazquez, I., Jonson, T., Yakowitz, D., Nair, R., and Hussein, S. 2015. "Great Expectations: The Evolution of the Chief Data Officer," http://www.pwc.com/fsi
Xu, Y., Li, Z., Gupta, A., Bugdayci, A., and Bhasin A. 2014. "Modeling Professional Similarity by Mining Professional Career Trajectories," in *Proceedings* of KDD'14, pages 1945-1954.
Yan, X., Guo, J., Liu, S., Cheng, X., and Wang, Y. 2013. "Learning Topics in Short Texts by Non-negative Matrix Factorization on Term Correlation Matrix," in *Proceedings* of the 2013 SIAM International Conference on Data Mining.
Xu, H., Yu, Z., Yang, J., Xiong, H., and Zhu, H. 2016. "Talent Circle Detection in Job Transition Networks," in *Proceedings* of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, Pages 655-664.
Xu, W., Liu, X., and Gong, Y. 2003. "Document Clustering Based on Non-negative Matrix Fctorization," in *Proceedings* of the 26th annual international ACM SIGIR Conference on Research and Development in Information Retrieval.